1   **Epigenome-wide meta-analysis of BMI in nine cohorts: examining the utility of epigenetic**
2   **BMI in predicting metabolic health**

3   Whitney L. Do[1], Dianjianyi Sun[2], Karlijn Meeks[3], Pierre-Antoine Dugue[4-6], Ellen
4   Demerath[7], Weihua Guan[8], Shengxu Li[9], Wei Chen[9], Roger Milne[4-6], Abedowale
5   Adeyemo[3], Charles Agyemang[10], Rami Nassir[11], JoAnn Manson[12], Aladdin H Shadyab[13],
6   Lifang Hou[14], Steve Horvath[15], Themistocles L. Assimes[16], Parveen Bhatti[17], Kristina
7   Jordahl[18], Andrea Baccarelli[19], Alicia Smith[20], Lisa R. Staimez[21], Aryeh Stein[21], Eric A.
8   Whitsel[22], K.M. Venkat Narayan*[21], Karen Conneely*[23]

9   [1]Laney Graduate School, Emory University, Atlanta, GA, USA, [2]School of Public Health and Tropical Medicine, Tulane University,
10   New Orleans, LA, USA, [3]Center for Research on Genomics and Global Health, National Human Genome Research Institute,
11   National Institutes of Health, Bethesda, MD, USA, [4]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia, ,
12   [5]Precision Medicine, School of Clinical Sciences At Monash Health, Monash University, Clayton, VIC, Australia, [6]Centre for
13   Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC,
14   3051, Australia[7]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN,
15   USA, [8]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA, [9]Children's Minnesota
16   Research Institute, Childrens Minnesota, Minneapolis, MN, USA, [10]Department of Public and Occupational Health, Amsterdam
17   Public Health Research Institute, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, the Netherlands,
18   [11]Department of Pathology, Umm Al-Qura University, Mecca, Saudi Arabia, [12]Department of Medicine, Brigham and Women's
19   Hospital, Harvard Medical School, Boston, MA, USA, [13]Herbert Wertheim School of Public Health and Human Longevity Science,
20   University of California, San Diego [14]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University,
21   Chicago, IL, USA, [15]Department of Human Genetics, University of California, Los Angeles, [16]Department of Medicine, School of
22   Medicine, Stanford, CA, USA, [17]Cancer Control Research, BC Cancer, Vancouver, BC, [18]Department of Epidemiology, University of
23   Washington, Seattle, WA, USA [19]Department of Environmental Health Sciences, Columbia University, New York City, NY, USA,
24   USA, [20]Department of Gynecology and Obstetrics, School of Medicine, Emory University, Atlanta, GA, USA, [21]Hubert Department of
25   Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA [22]Departments of Epidemiology and Medicine,
26   University of North Carolina, Chapel Hill, NC, [23]Department of Human Genetics, Emory University, Atlanta, GA, USA

27   *Contributed equally

28   Correspondence: Whitney L. Do, whitney.leet@nih.gov, Karen N. Conneely, kconnee@emory.edu

29   **Abstract**

30   This study sought to examine the association between DNA methylation and body mass index

31   (BMI) and the potential utility of these cytosine-phosphate-guanine (CpG) sites in predicting

32   metabolic health. We pooled summary statistics from six trans-ethnic EWAS of BMI

33   representing nine cohorts (n=17058), replicated these findings in the Women's Health Initiative

34   (WHI, n=4822) and developed an epigenetic prediction score of BMI. In the pooled EWAS, 1265

35   CpG sites were associated with BMI (p<1E-7), and 1238 replicated in the WHI (FDR < 0.05).

36   We performed several stratified analyses to examine whether these associations differed

37   between individuals of European descent and individuals of African descent. We found five CpG

38   sites had a significant interaction with BMI by race/ethnicity. To examine the utility of the

39   significant CpG sites in predicting BMI, we used elastic net regression to predict log normalized

40   BMI in the WHI (80% training/20% testing). This model found 397 sites could explain 32% of the

41   variance in BMI in the WHI test set. Individuals whose methylome-predicted BMI overestimated

42   their BMI (high epigenetic BMI) had significantly higher glucose and triglycerides, and lower

43   HDL-cholesterol and LDL-cholesterol compared to accurately predicted BMI. Individuals whose

44   methylome-predicted BMI underestimated their BMI (low epigenetic BMI) had significantly

45   higher HDL-cholesterol and lower glucose and triglycerides. This study identified 553 previously

46   identified and 685 novel CpG sites associated with BMI. Participants with high epigenetic BMI

47   had poorer metabolic health suggesting that the overestimation may be driven in part by

48   cardiometabolic derangements characteristic of metabolic syndrome.

49
50
51
52
53

Introduction

54

55   Globally, the prevalence of obesity is rising with an estimated 650 million adults obese,

56   representing 19.5% of the adult population [1, 2]. Obesity has been found to accompany a

57   multitude of molecular and metabolic perturbations including impaired cell signaling, insulin

58   resistance, hyperlipidemia, and hypertension [3-5]. Ultimately these perturbations can lead to

59   the early onset of chronic diseases with individual living with obesity having a 37% increased

60   risk of type 2 diabetes [6] and 67-85% increased risk of cardiovascular disease compared to

61   individuals living without obesity [7]. With a growing population of individuals living with obesity,

62   it is increasingly important to understand the molecular mechanisms dysregulated by obesity to

63   further elucidate both early markers of disease progression and novel therapeutic targets.

64   Epigenetic mechanisms are molecularly-mediated changes in gene function which do

65   not change the DNA sequence. DNA methylation, the most widely characterized epigenetic

66   mechanism, occurs when a methyl group attaches to the cytosine in a cytosine-guanine

67   nucleotide (CpG) pair [8]. DNA methylation has been shown to influence gene expression by

68   blocking transcription factor binding and recruiting chromatin remodelers [9]. As a functional

69   mechanism influencing gene expression, DNA methylation may be on a disease pathway and

70   could provide insight into important therapeutic targets. DNA methylation has also become an

71   important biomarker of health, for example with the development of epigenetic clocks, which can

72   provide accurate estimates of individual age based on the methylation status of a representative

73   set of CpG sites [10]. Individuals whose DNA methylation deviate from their actual chronological

74   age, such that their epigenetically predicted age is higher than their actual age, have been

75   shown to have higher rates of cancer, cardiovascular disease, diabetes, and mortality [11]. All of

76   these properties may be relevant in the relationship between DNA methylation and obesity.

77   Several studies have examined the relationship between DNA methylation and body

78   mass index (BMI), a commonly used measure of obesity [12-21]. Obesity has been significantly

79   associated with differential DNA methylation, and Mendelian randomization analyses have

80   suggested that while this differential methylation appears to be a consequence of the state of

81   obesity at many CpG sites, some CpGs show evidence consistent with causal roles in obesity

82   [13, 18]. While several large-scale studies have identified sites associated with obesity, it is

83   likely that additional sites will be detectable only with large sample sizes, as observed DNA

84   methylation differences are often subtle [22]. Thus, a goal of this study is to conduct the largest

85   epigenome-wide association study (EWAS) meta-analysis of BMI in nine population-based

86   cohort studies to identify novel sites associated with obesity. The identification of novel sites can

87   reveal unique molecular signatures of various BMI phenotypes (including metabolically

88   healthy/unhealthy BMI) and may enable improved prediction of BMI. Previous studies have

89   reported that a collection of methylation-based predictors can explain between 4.7-18% of the

90   variance in BMI [13, 21, 23, 24]. In conducting the largest EWAS, we may have better predictive

91   capacity by incorporating the novel CpG sites identified in the EWAS meta-analysis. As such, a

92   secondary aim of this study is to examine whether BMI-associated CpG sites can predict BMI.

93   As with epigenetic age, deviations from epigenetically predicted BMI may be associated with

94   several relevant health outcomes and could be used as an informative metric of overall health

95   and/or a predictor of future cardiovascular disease. Thus, we examined whether individuals

96   whose BMI was poorly predicted by DNA methylation (DNA methylation over predicts their

97   actual BMI or DNA methylation under predicts their actual BMI) have differential metabolic

98   health status.

99   **Methods**

100   *Participants*

101   Our discovery analysis used data from 17,034 participants from six published EWAS

102   studies of individuals of European descent (n=11220), African descent (n=2587), and South

103   Asian descent (n=2680). The six studies were based on nine cohorts: Atherosclerosis Risk in

4

104   Communities study (ARIC) [25], Melbourne Collaborative Cohort Study (MCCS) [26], Lifelines

105   DEEP [27], Lothian Birth Cohort (LBC) 1921 and 1936 [28], Bogalusa Heart Study (BHS) [29],

106   the Research on Obesity and Diabetes among African Migrants (RODAM) [30], the Kooperative

107   Gesundheitsforschung in der Region Augsburg (KORA) [31], the London Life Sciences

108   Prospective Population Study (LOLIPOP) [32], and Italian cardiovascular component of the

109   European Prospective Investigation into Cancer and Nutrition (EPICORE) [33]. Replication

110   analyses were conducted in three ancillary studies from the Women's Health Initiative (WHI):

111   *Epigenetic Mechanisms of Particulate Matter-Mediated Cardiovascular Disease* (EMPC, aka

112   AS315), the *Integrative Genomics for Risk of Coronary Heart Disease and Related Phenotypes*

113   *in WHI cohort* (BAA23), and *Bladder Cancer and Leukocyte Methylation* (AS311). In the WHI,

114   individuals were excluded if BMI and blood samples for DNA methylation were not measured

115   within the same year. Extreme levels of BMI <17 kg/m^2 and >75 kg/m^2 were excluded. Further

116   description of the discovery and replication cohorts is described in **Supplemental Methods.**

117   *BMI, DNA methylation and covariates*

118   BMI was defined as weight in kg/height in m^2. Methodologies obtaining weight and

119   height differed among the studies, however all used standard methods. One study transformed

120   BMI values to obtain a normal distribution [20]. Relevant variables in our replication analysis

121   included race/ethnicity, age, physical activity and smoking status. Race/ethnicity, smoking and

122   physical activity were self-reported. Smoking status was defined as current, former or never.

123   DNA methylation was measured in several cell types including CD4+ T-cells,

124   mononuclear cells and whole blood. DNA methylation in all studies was measured using the

125   Illumina 450K Infinium Methylation BeadChip. DNA methylation was estimated as the proportion

126   of methylated signal relative to combined unmethylated and methylated signal for a specific

127   CpG site, defined as the β-value. Quality control procedures of the previous studies have been

128   reported in detail and they did not differ substantially across studies. In the WHI, all methylation

129   data were quality controlled and normalized using beta-mixture quantile normalization. In

130   replication analyses, chip and row were included as technical covariates in all models to adjust

131   for batch effects. Cell composition was estimated using methods derived by Houseman et al.

132   [34].

133   *Statistical Analysis*

134   A summary of our analyses is included in **Figure 1A**. Our primary method was weighted

135   sum of Z-score meta-analysis [35]. This method utilizes Z-scores from individual study summary

136   statistics computed from inverse-normal p-values and the direction of effect to determine

137   significant sites. This was chosen as the primary method for meta-analysis since the studies did

138   not all have equivalent exposure-outcome definition (DNA methylation defined as exposure in

139   two studies and outcome in four studies) and BMI was transformed in one study. The EWAS

140   was adjusted for genomic inflation and significance was defined as $p < 1\times10^{-7}$.

141   The significant sites were examined for replication within WHI. Models were stratified by

142   ancillary study. Covariates in this analysis included age, race/ethnicity, cell composition, the top

143   three principal components of genetic relatedness, smoking status, clinical trial arm and case-

144   control status (BAA23 and AS311). To account for potential chip-to-chip differences in

145   measurement and to adjust for batch effects, chip was included as a random effect for each

146   BeadChip in our model. Stratified analyses were combined using inverse-variance weighted

147   (IVW) meta-analysis [36]. Significance was defined by false discovery rate (FDR) q-value <

148   0.05.

149   *BMI Prediction Score*

150   To examine the degree to which methylation can predict BMI and the secondary

151   cardiometabolic outcomes associated with BMI, we used elastic net regression models with the

152   significant sites to predict log-normalized BMI. The WHI cohorts were randomly divided into a

153   training and test set (80% and 20%, respectively) with an equal BMI distribution. We used

154   elastic net regression on the training set with 10-fold cross validation to select a predictive

155   model, which we subsequently tested in the test set. Using the significant sites and coefficients

156   selected by the model, a DNA methylation prediction score was developed by multiplying the

157   coefficient by the individual β-value and summing over all the sites for each individual. We then

158   evaluated the performance of the DNA methylation score in the test set, both in terms of how

159   accurately it predicted BMI (metrics: R^2 and median absolute deviation) and how well it

160   predicted obesity status (BMI ≥ 30 kg/m^2) (metrics: sensitivity and specificity).

161   Using the predicted BMI values, we examined the patterns among outliers in the prediction

162   model. Individuals were split into categories based on regressing the predicted BMI on the

163   actual BMI. Accurately predicted individuals were defined as those with residuals between -0.04

164   to 0.04 (accurate epigenetic BMI). Individuals outside of this range were split into two groups:

165   residual below -0.04 (low epigenetic BMI or individuals whose methylome-predicted BMI

166   underestimated their BMI) and residual above 0.04 (high epigenetic BMI or individuals whose

167   methylome-predicted BMI overestimated their BMI). These thresholds were defined based on

168   the 10% and the 90% distribution of the residuals. Using these categories, we examined

169   cardiometabolic differences including waist circumference, triglycerides, HDL-cholesterol, LDL-

170   cholesterol, and blood glucose among these categories using linear regression models

171   regressing log-normalized cardiometabolic markers on DNA methylation prediction category

172   adjusted for age, race/ethnicity, smoking status and physical activity. To aid interpretability,

173   results were reported based on the change in average value in the text. We additionally

174   examined results using thresholds defined by the 20% and 80% distribution of residuals and

175   found consistent findings.

176   *Sensitivity Analyses*

7

177   We conducted several sensitivity analyses. In the discovery meta-analysis, we examined

178   the influence of specific studies on the results using leave-one out analyses to examine the

179   degree that each study is influencing the results. We also compared results from the weighted

180   sum of z-score meta-analysis with results that would be obtained using an IVW meta-analysis in

181   studies with the same exposure-outcome definition. We examined several sites for interaction

182   by self-reported race/ethnicity and BMI using linear mixed-effect models adjusting for age, cell

183   composition, smoking status, WHI study randomization arm, case-control status, row with a

184   random effect for chip. In our replication analysis in WHI, models were additionally adjusted for

185   diet quality, physical activity level and socioeconomic status.

186   **Results**

187   Our discovery analysis included 17058 participants from six EWAS (**Figure 1A**, **Table 1**

188   and **Supplementary Table 1**). The definition of BMI and DNA methylation differed with several

189   transforming these values in the models (**Table 1**). The covariates in the model also differed

190   with all studies adjusting for age and sex, and the majority adjusting for cell composition and

191   smoking status. When pooling results from all studies, 1265 CpG sites were associated with

192   BMI (**Figure 1B**, **Supplementary Table 2**, $p < 1E-7$) with 498 of the sites having a consistent

193   direction of effect in all of the cohorts meta-analyzed. More than half of the significant sites (726

194   CpG sites) were positively associated with BMI.

195   In the WHI, 367 women were excluded due to missing BMI, extreme levels of BMI, or

196   overlap leaving 4822 women included in the replication cohort (**Supplementary Table 3**). Of

197   the 1265 sites identified in the discovery analysis, 1254 were analyzed after QC. In the WHI,

198   1238 CpG sites were significantly associated with BMI (**Supplementary Table 4**, FDR q-value <

199   0.05). These 1238 CpG sites annotated to 742 unique genes. Additionally, 147 of these genes

200   were annotated to more than one BMI-associated CpG site, with 382 CpG sites annotated to

201   these 147 genes. With the large sample size, we were able to discover 685 novel CpG sites that

202 had not previously been identified in EWAS of BMI as well as 553 CpG sites previously

203 identified in the literature. We examined how the replicated sites associated with differential

204 gene expression based on previously published analyses of the Grady Trauma Project (GTP)

205 and Multi-Ethnic Study of Atherosclerosis (MESA) cohort [37]. The 1238 CpG sites associated

206 with 1103 CpG-mRNA associations in MESA (**Supplementary Table 5**) and 79 CpG-mRNA

207 associations in GTP (**Supplementary Table 6**). One site associated with the same mRNA

208 transcript in both cohorts, cg25653947, which was positively associated with expression in

209 *TOP1MT*. We performed a gene ontology (GO) analysis of the differentially expressed genes

210 and found enrichment in pathways related to the adaptive immune system with regulation in B-

211 and T-cell pathways (**Figure 1C**, **Supplementary Table 7**).

212 We next re-performed our discovery EWAS stratified by European vs. African descent.

213 We found 936 and 130 CpG sites that were associated with BMI in the analyses restricted to

214 individuals from European (n=11,220) and African (n=2,587) descent, respectively. Of the 130

215 significant CpG sites in the analysis of individuals of African descent, 43 unique sites were only

216 significant in that population (**Supplementary Table 8-9**). We examined these sites for

217 interaction in the WHI non-Hispanic white and African American individuals. We found that five

218 CpG sites had a significant interaction with BMI by race/ethnicity (**Table 2, Supplementary**

219 **Fig.1**). Two sites were quantitative trait methylation loci in the GTP cohort: cg25212453

220 negatively associated with *TNFRSF13B* and *COCH* and cg08122652 negatively associated with

221 *LGALS3BP* and *OTOF* (**Supplementary Table 10**).

222 We next explored the potential of DNA methylation to predict BMI using the 1238 CpG

223 sites from the replication analysis. After model tuning using elastic-net regression in a training

224 set (N=3858), 398 sites were selected for the model (**Supplementary Table 11**). These sites

225 accounted for 32% of the variance in BMI in the test set (MAD = 0.04, N=964). The addition of

226 age, race/ethnicity, physical activity, and cell composition as predictors only marginally

227 improved the adjusted R^2 (**Table 4**). In the combined training and test set (N=4822), these sites

228 accounted for 36% of the variance in BMI. For comparison we examined how well the predictors

229 from Mendelson *et al.* [13] estimated BMI in the WHI cohort. In the full WHI cohort, the 83 CpG

230 sites accounted for 29% of the variance.

231   We next assessed the potential of this DNA methylation-based BMI score to predict

232 obesity, defined as BMI>30. In our test set (N=964), the sensitivity was 0.82 and the specificity

233 was 0.57 with an area under the curve of 0.69 (**Figure 2A**). Individuals were then categorized

234 based on how well methylation predicted BMI. On average, DNA methylation tended to

235 underpredict BMI (**Figure 2B**). Individuals with high epigenetic BMI had 20.5 mg/dL higher

236 blood glucose (SE: 2.0, p<2E-16), 31 mg/dL higher triglycerides (SE: 4.3, p=9.24E-08), 4.3

237 mg/dL lower HDL-cholesterol (SE: 0.68, p=1.06E-07) and 3.3 mg/dL lower LDL-cholesterol (SE:

238 2.0, p=0.047) compared with accurate predicted BMI. In contrast, individuals with low epigenetic

239 BMI had 5.2 mg/dL lower blood glucose (SE: 2.2, p<2E-16), 23.7 mg/dL lower triglycerides (SE:

240 4.8, p=2.39E-08) and 3.0 mg/dL higher HDL-cholesterol (**Figure 2C**, SE: 0.8, p=0.0004)

241 compared to accurate predicted BMI**.**

242   We conducted several sensitivity analyses. We first examined how the results changed

243 in a leave-one out meta-analysis (**Supplementary Table 12**). Excluding the results from Wahl

244 *et al.* [18] led to the largest reduction in significant sites resulting in 536 significant CpG sites.

245 However, this is likely due to a reduction in power. We next compared results obtained using Z-

246 score vs. IVW meta-analysis in cohorts with the same exposure-outcome relationship (ARIC,

247 RODAM, BHS White and BHS Black). In the IVW and Z-score meta-analysis of these four

248 studies, 1939 CpG sites and 1433 CpG sites, respectively, were significantly associated with

249 BMI (p < 1E-7) with 935 overlapping sites among methods. Among the sites identified significant

250 in either analysis, the correlation between the test statistics obtained using Z-score vs. IVW

251 meta-analysis was 0.98. The meta test statistics tend to be smaller when identified using

252   weighted sum of Z-score meta-analysis, suggesting our main results may be more conservative

253   than what would be obtained using IVW meta-analysis. Finally, in the main EWAS, we

254   examined how the results changed when adjusted for diet, physical activity and income status.

255   Overall, 1161, 1167 and 1160 CpG sites remained associated with BMI when additional

256   covariates were included to adjust for diet quality, physical activity and income, respectively.

257   **Discussion**

258   This study identified a unique methylomic signature of BMI and obesity. In the WHI, the

259   majority of the sites identified in the discovery cohort (99%) were replicated and found to predict

260   several metabolic and inflammatory pathways. Moreover, we found five CpG sites that are

261   differentially associated with BMI between non-Hispanic whites and African Americans, two of

262   which may play a role in gene expression. Finally, we constructed a score based on 398 CpG

263   sites that was able to predict BMI as well as several other cardiometabolic risk factors.

264   Individuals whose measured BMI was higher than predicted by their methylome were found to

265   have poorer metabolic health including higher blood glucose and triglycerides and lower HDL-

266   cholesterol compared to individuals whose BMI was accurately predicted.

267   This study identified 1238 CpG sites that were significantly associated with BMI in

268   several race/ethnicity groups. The 1238 CpG sites were associated with differential gene

269   expression in MESA and GTP. In the GO analysis of the differentially expressed transcripts, the

270   most significant enriched pathways were immune response pathways, particularly the adaptive

271   immune response. The top pathways regulated B- and T-cell signaling. Low-grade inflammation

272   in obesity is a hallmark of the disease, which leads to significant metabolic dysregulation [38].

273   Several studies have found BMI-associated CpG sites are enriched for immune pathways [13,

274   39]. Mechanistic studies have identified DNA methylation as playing a key role in promoting

275   macrophage polarization in response to obesity, with more M1 macrophages associated with

276   obesity [40].

11

277   We were also able to examine how these associations differed when stratified by

278   race/ethnicity. Racial and ethnic differences in adiposity have been well established. While

279   African Americans have been found to have higher risk for cardiovascular diseases compared to

280   non-Hispanic whites, they have consistently been shown to have lower visceral adipose tissue

281   and lower body fat percentage compared to non-Hispanic whites [41, 42]. Of the five sites with a

282   significant interaction between BMI and race/ethnicity, two CpG sites were associated with

283   differential expression in four mRNA transcripts. These gene transcripts are related to

284   inflammatory pathways and hearing. *TNFRSF13B* and *LGALS3BP* were differentially expressed

285   in association with two CpG sites. These two genes have been found to be regulators of NF-

286   kappa-B signaling and to be upregulated with obesity [43-45]. Our study found a positive

287   association between BMI and methylation in cg25212453 and cg08122652 (in WHI) and a

288   negative association between methylation in these two sites and expression in *LGALS3BP* and

289   *TNFRSF13B* (in GTP). Thus, as BMI increases in African Americans, gene expression may be

290   decreasing in these sites, suggesting a potentially advantageous effect on inflammatory profiles

291   in African Americans. Low-grade inflammation in obesity leads to significant metabolic

292   dysregulation [38]. However, there is some epidemiological data that suggests individuals of

293   African descent may not be as prone to an increased inflammatory profiles when living with

294   obesity [46, 47]. Our study may provide some mechanistic explanation to these differences in

295   the relationship between inflammation and adiposity in individuals of African descent.

296   We also found that DNA methylation was predictive of BMI, with the score we developed

297   based on 398 CpG sites explaining 32% of the variance in BMI in an independent test set.

298   Previous studies constructing scores based on smaller samples have been able to explain

299   between 4.7-18% of the variance in BMI [13, 21, 23, 24]. DNA methylation has been found to be

300   an accurate predictor of current BMI and a poor predictor of future BMI [24]. Outliers in the

301   epigenetic BMI model predicted a unique phenotype. Individuals with high epigenetic BMI or

302   whose BMI was over-predicted by the epigenetic markers had poorer cardiometabolic markers

303   compared to accurately predicted BMI. This may suggest that epigenetic BMI prediction may be

304   identifying individuals with poor health regardless of their BMI and that these sites may be

305   useful biomarkers to examine further. Our findings related to LDL-C were inconsistent with other

306   cardiometabolic markers. We found that individuals with high epigenetic BMI had lower LDL-C

307   as compared to individuals with accurate epigenetic BMI. These results were verging on null

308   (p=0.0497), thus they should be interpreted with caution.

309   Some of our findings should be interpreted with caution given several important

310   limitations. In the discovery analysis, we stratified this analysis based on race/ethnicity as it was

311   defined within each of the individual studies. This differed between studies with most based on

312   self-report of race/ethnicity. Thus, it is unclear whether we are identifying molecular differences

313   due to ancestry or social construct. Moreover, these populations, which include African

314   Americans, Ghanaians, and European-residing Ghanaians, are not homogenous in genetic

315   ancestry, living environment, lifestyles and other factors. Nevertheless, our interaction and

316   expression analyses were conducted in African American populations from the WHI and GTP,

317   so these results may only be generalizable for this population. In particular, the racial disparities

318   in the US may be an underlying cause of these results, as opposed to differences in ancestry.

319   For example in the US, African Americans are much more likely to live in poverty compared to

320   non-Hispanic whites [48]. In our results, we may be identifying compensatory mechanisms of

321   structural racism which may be driven by environmental exposures for example, ambient

322   particulate matter exposure, stress, lack of access to health care as well as obesity. Another

323   potential limitation is that the training and test set in our prediction analyses come from the

324   same population (WHI). Future research efforts could test this model in another population to

325   examine the reproducibility of these findings.

326   Overall, this study yields several important discoveries. We identified novel sites

327   associated with BMI and found a unique molecular profile associated with obesity in individuals

328   of African descent. We additionally found that epigenetic markers can predict BMI well and may

329   be able to distinguish individuals whose metabolic health does not align with their BMI. Future

330   studies should examine whether BMI-associated methylation is differential by metabolic health

331   status.

332   ***Declaration of Interests***

333   KMJ is an employee of Bristol Myers Squibb.

334   ***Acknowledgments***

379   ***Author Contributions***

380   WLD, KVN, and KNC conceived of the study. WLD, KVN, LRS, ADS and AS developed the

381   methods. ED, WG, SL, WC, DS, PMV, SS, RM, PAD, KM, AA, and CA computed and shared

382   the summary statistics from the participating studies in the discovery analysis. LH, SH, TA, PB,

383   KMJ, EAW and AB curated and processed the data used in the replication analyses. WLD

384   performed all statistical analyses with support from KNC. WLD drafted the manuscript with

385   content expertise and review from all authors. All authors read and approved of the final

386   manuscript.

387   ***Data and Code Availability***

388   All datasets used in this study are publicly available at individual study websites. The code used

389   to generate these findings are available from the corresponding author upon reasonable

390   request.

References

1. Organisation for Economic Co-operation and Development: **Obesity update 2017**. In.: OECD; 2017.
2. World Health Organization: **Obesity and Overweight**. In.: WHO; 2016.
3. Asghar A, Sheikh N: **Role of immune cells in obesity induced low grade inflammation and insulin resistance**. *Cell Immunol* 2017, **315**:18-26.
4. O'Flanagan CH, Bowers LW, Allot EH, Hursting SD: **Molecular and metabolic mechanisms underlying the obesity–cancer link**. In: *Energy Balance and Obesity.* vol. 10: IARC Working Group; 2017.
5. Hall JE, Carmo JMd, Silva AAd, Wang Z, Hall ME: **Obesity-Induced Hypertension**. *Circulation research* 2015, **116**(6):991-1006.
6. Narayan KMV, Boyle JP, Thompson TJ, Gregg EW, Williamson DF: **Effect of BMI on Lifetime Risk for Diabetes in the U.S**. *Diabetes care* 2007, **30**(6):1562-1566.
7. Khan SS, Ning H, Wilkins JT, Allen N, Carnethon M, Berry JD, Sweis RN, Lloyd-Jones DM: **Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity**. *JAMA cardiology* 2018, **3**(4):280-287.
8. Smith ZD, Meissner A: **DNA methylation: roles in mammalian development**. *Nature reviews Genetics* 2013, **14**(3):204-220.
9. Greenberg MVC, Bourc'his D: **The diverse roles of DNA methylation in mammalian development and disease**. *Nature Reviews Molecular Cell Biology* 2019, **20**(10):590-607.
10. Horvath S, Raj K: **DNA methylation-based biomarkers and the epigenetic clock theory of ageing**. *Nature Reviews Genetics* 2018, **19**(6):371-384.
11. Fransquet PD, Wrigglesworth J, Woods RL, Ernst ME, Ryan J: **The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis**. *Clinical epigenetics* 2019, **11**(1):62.
12. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou Y-H, Hedman ÅK, Sandling JK, Li L-A, Irvin MR *et al*: **Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci**. *Human Molecular Genetics* 2015, **24**(15):4464-4479.
13. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, Demerath EW, Guan W, Zhi D, Yao C *et al*: **Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach**. *PLoS medicine* 2017, **14**(1):e1002215.
14. Al Muftah WA, Al-Shafai M, Zaghlool SB, Visconti A, Tsai PC, Kumar P, Spector T, Bell J, Falchi M, Suhre K: **Epigenetic associations of type 2 diabetes and BMI in an Arab population**. *Clinical epigenetics* 2016, **8**:13.
15. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, Sha J, Pankow JS, Liu C, Irvin MR *et al*: **Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference**. *Obesity (Silver Spring, Md)* 2015, **23**(7):1493-1501.
16. Campanella G, Gunter MJ, Polidoro S, Krogh V, Palli D, Panico S, Sacerdote C, Tumino R, Fiorito G, Guarrera S *et al*: **Epigenome-wide association study of adiposity and future risk of obesity-related diseases**. *International journal of obesity (2005)* 2018, **42**(12):2022-2035.
17. Meeks KAC, Henneman P, Venema A, Burr T, Galbete C, Danquah I, Schulze MB, Mockenhaupt FP, Owusu-Dabo E, Rotimi CN *et al*: **An epigenome-wide association study in whole blood of**

436   measures of adiposity among Ghanaians: the RODAM study. *Clinical epigenetics* 2017, **9**:103-
437   103.
438 18. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai P-C, Ried JS, Zhang W, Yang Y *et al*:
439   **Epigenome-wide association study of body mass index, and the adverse outcomes of**
440   **adiposity**. *Nature* 2017, **541**(7635):81-86.
441 19. Sun D, Zhang T, Su S, Hao G, Chen T, Li Q-Z, Bazzano L, He J, Wang X, Li S *et al*: **Body Mass Index**
442   **Drives Changes in DNA Methylation**. *Circulation research* 2019, **125**(9):824-833.
443 20. Geurts YM, Dugue PA, Joo JE, Makalic E, Jung CH, Guan W, Nguyen S, Grove ML, Wong EM,
444   Hodge AM *et al*: **Novel associations between blood DNA methylation and body mass index in**
445   **middle-aged and older adults**. *International journal of obesity (2005)* 2018, **42**(4):887-896.
446 21. Shah S, Bonder Marc J, Marioni Riccardo E, Zhu Z, McRae Allan F, Zhernakova A, Harris Sarah E,
447   Liewald D, Henders Anjali K, Mendelson Michael M *et al*: **Improving Phenotypic Prediction by**
448   **Combining Genetic and Epigenetic Associations**. *The American Journal of Human Genetics* 2015,
449   **97**(1):75-85.
450 22. Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, Herbstman J, Holland N,
451   LaSalle JM, Schmidt R *et al*: **Small-Magnitude Effect Sizes in Epigenetic End Points are**
452   **Important in Children's Environmental Health Studies: The Children's Environmental Health**
453   **and Disease Prevention Research Center's Epigenetics Working Group**. *Environ Health Perspect*
454   2017, **125**(4):511-526.
455 23. McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, Morris SW,
456   Bermingham ML, Campbell A, Murray AD *et al*: **Epigenetic prediction of complex traits and**
457   **death**. *Genome Biol* 2018, **19**(1):136.
458 24. Reed ZE, Suderman MJ, Relton CL, Davis OSP, Hemani G: **The association of DNA methylation**
459   **with body mass index: distinguishing between predictors and biomarkers**. *Clinical epigenetics*
460   2020, **12**(1):50.
461 25. **The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC**
462   **investigators**. *Am J Epidemiol* 1989, **129**(4):687-702.
463 26. Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK, Bruinsma FJ, Lynch BM,
464   Dugué PA, Jayasekara H *et al*: **Cohort Profile: The Melbourne Collaborative Cohort Study**
465   **(Health 2020)**. *International Journal of Epidemiology* 2017, **46**(6):1757-1757i.
466 27. Tigchelaar EF, Zhernakova A, Dekens JAM, Hermes G, Baranska A, Mujagic Z, Swertz MA, Muñoz
467   AM, Deelen P, Cénit MC *et al*: **Cohort profile: LifeLines DEEP, a prospective, general population**
468   **cohort study in the northern Netherlands: study design and baseline characteristics**. *BMJ Open*
469   2015, **5**(8):e006772.
470 28. Deary IJ, Gow AJ, Taylor MD, Corley J, Brett C, Wilson V, Campbell H, Whalley LJ, Visscher PM,
471   Porteous DJ *et al*: **The Lothian Birth Cohort 1936: a study to examine influences on cognitive**
472   **ageing from age 11 to age 70 and beyond**. *BMC Geriatr* 2007, **7**:28-28.
473 29. Berenson GS: **Bogalusa Heart Study: a long-term community study of a rural biracial**
474   **(Black/White) population**. *Am J Med Sci* 2001, **322**(5):293-300.
475 30. Agyemang C, Beune E, Meeks K, Owusu-Dabo E, Agyei-Baffour P, Aikins A, Dodoo F, Smeeth L,
476   Addo J, Mockenhaupt FP *et al*: **Rationale and cross-sectional study design of the Research on**
477   **Obesity and type 2 Diabetes among African Migrants: the RODAM study**. *BMJ Open* 2014,
478   **4**(3):e004877.
479 31. Holle R, Happich M, Löwel H, Wichmann HE: **KORA--a research platform for population based**
480   **health research**. *Gesundheitswesen* 2005, **67 Suppl 1**:S19-25.
481 32. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, Wahl S, Elliott HR, Rota F, Scott WR *et*
482   *al*: **Epigenome-wide association of DNA methylation markers in peripheral blood from Indian**

483   **Asians and Europeans with incident type 2 diabetes: a nested case-control study**. *Lancet*
484   *Diabetes Endocrinol* 2015, **3**(7):526-534.

485 33. Bueno H, Danchin N, Tafalla M, Bernaud C, Annemans L, Van de Werf F: **EPICOR (long-tErm**
486   **follow-up of antithrombotic management Patterns In acute CORonary syndrome patients)**
487   **study: rationale, design, and baseline characteristics**. *Am Heart J* 2013, **165**(1):8-14.

488 34. Houseman EA, Molitor J, Marsit CJ: **Reference-free cell mixture adjustments in analysis of DNA**
489   **methylation data**. *Bioinformatics (Oxford, England)* 2014, **30**(10):1431-1439.

490 35. Lee CH, Cook S, Lee JS, Han B: **Comparison of Two Meta-Analysis Methods: Inverse-Variance-**
491   **Weighted Average and Weighted Sum of Z-Scores**. *Genomics Inform* 2016, **14**(4):173-180.

492 36. Woolf B: **On estimating the relation between blood group and disease**. *Ann Hum Genet* 1955,
493   **19**(4):251-253.

494 37. Kennedy EM, Goehring GN, Nichols MH, Robins C, Mehta D, Klengel T, Eskin E, Smith AK,
495   Conneely KN: **An integrated -omics analysis of the epigenetic landscape of gene expression in**
496   **human blood cells**. *BMC Genomics* 2018, **19**(1):476.

497 38. Reilly SM, Saltiel AR: **Adapting to obesity with adipose tissue inflammation**. *Nature Reviews*
498   *Endocrinology* 2017, **13**(11):633-643.

499 39. Do WL, Nguyen S, Yao J, Guo X, Whitsel EA, Demerath E, Rotter JI, Rich SS, Lange L, Ding J *et al*:
500   **Associations between DNA methylation and BMI vary by metabolic health status: a potential**
501   **link to disparate cardiovascular outcomes**. *Clinical Epigenetics* 2021, **13**(1):230.

502 40. Wang X, Cao Q, Yu L, Shi H, Xue B, Shi H: **Epigenetic regulation of macrophage polarization and**
503   **inflammation by DNA methylation in obesity**. *JCI Insight* 2016, **1**(19):e87748-e87748.

504 41. Després J-P: **Body Fat Distribution and Risk of Cardiovascular Disease**. *Circulation* 2012,
505   **126**(10):1301-1313.

506 42. Dugas LR, Cao G, Luke AH, Durazo-Arvizu RA: **Adiposity is not equal in a multi-race/ethnic**
507   **adolescent population: NHANES 1999-2004**. *Obesity (Silver Spring)* 2011, **19**(10):2099-2101.

508 43. Zhang Y, Kent Jr. JW, Olivier M, Ali O, Broeckel U, Abdou RM, Dyer TD, Comuzzie A, Curran JE,
509   Carless MA *et al*: **QTL-based association analyses reveal novel genes influencing pleiotropy of**
510   **metabolic syndrome (MetS)**. *Obesity* 2013, **21**(10):2099-2111.

511 44. Xu G, Xia Z, Deng F, Liu L, Wang Q, Yu Y, Wang F, Zhu C, Liu W, Cheng Z *et al*: **Inducible**
512   **LGALS3BP/90K activates antiviral innate immune responses by targeting TRAF6 and TRAF3**
513   **complex**. *PLOS Pathogens* 2019, **15**(8):e1008002.

514 45. Challa TD, Straub LG, Balaz M, Kiehlmann E, Donze O, Rudofsky G, Ukropec J, Ukropcova B,
515   Wolfrum C: **Regulation of De Novo Adipocyte Differentiation Through Cross Talk Between**
516   **Adipocytes and Preadipocytes**. *Diabetes* 2015, **64**(12):4075-4087.

517 46. Doumatey AP, Lashley KS, Huang H, Zhou J, Chen G, Amoah A, Agyenim-Boateng K, Oli J,
518   Fasanmade O, Adebamowo CA *et al*: **Relationships Among Obesity, Inflammation, and Insulin**
519   **Resistance in African Americans and West Africans**. *Obesity* 2010, **18**(3):598-603.

520 47. Hakim O, Bello O, Ladwa M, Peacock JL, Umpleby AM, Charles-Edwards G, Amiel SA, Goff LM:
521   **The Link between Obesity and Inflammatory Markers in the Development of Type 2 Diabetes**
522   **in Men of Black African and White European Ethnicity**. *Nutrients* 2020, **12**(12):3796.

523 48. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CAM, Bertoni AG, Mujahid MS,
524   Palaniappan L, Taylor HA, Willis M *et al*: **Cardiovascular Health in African Americans: A**
525   **Scientific Statement From the American Heart Association**. *Circulation* 2017, **136**(21):e393-
526   e423.

527

528   ***Figure Legends***

529   Figure 1.

530   1A Description of study analyses. 1B Manhattan plot of the association between DNA

531   methylation and BMI. 1C Top pathways identified in gene ontology analysis.

532

533   Figure 2.

534   2A Receiver operating characteristic curve showing the performance of the DNA methylation

535   prediction score identifying obesity. AUC denotes area under the curve. Y-axis is the sensitivity

536   (true positive rate) and the x-axis is 1-specificity (false positive rate). 2B Scatter plot of predicted

537   BMI from elastic net regression of 398 CpG sites by actual BMI. Individuals categorized based

538   on the residual of predicted BMI regressed on actual BMI. 2C Boxplot of the association

539   between epigenetic prediction category and blood glucose (mg/dL), high density lipoprotein

540   (HDL-C, mg/dL), low density lipoprotein (LDL-C, mg/dL) and triglycerides (mg/dL).

541

542

Tables

Table 1. *Study characteristics of discovery analyses.*

Reference	Study Population	N	Exposure	Outcome	Sample	Covariates
Demerath, et al., 2015, Hum Mol Genet	Atherosclerosis Risk in Communities (ARIC)	2097	BMI	DNA methylation β-value	Leucocytes	Age, sex, study center, total white blood cell differentials, education, household income, cigarette smoking, current alcohol use, leisure physical activity, cell composition (Housman), top 10 PCs of genetic relatedness and batch effects (row, plate number and chip number)
Geurts, et al., 2018, Int J Obes	Melbourne Collaborative Cohort Study (MCCS)	5361	BMI Z-score	DNA methylation M-values	Dried blood spot, mononuclear cells, buffy coats	Age, sex, smoking status, country of birth, sample type, cell composition (Houseman), and study, plate and chip included as random effects.
Meeks, et al., 2017, Clin Epigenetics	Research on Obesity and Diabetes among African Migrants (RODAM) study	547	BMI	DNA methylation M-values	Whole Blood	Age, sex, recruitment site, cell composition (Houseman), hybridization batch, array position and first PC of genetic relatedness

Shah, et al., 2015, Am J Hum Genet	Lothian Birth Cohort (LBC) and Lifelines DEEP	2116	Log-transformed DNA methylation	BMI Z-score	Whole Blood	Age, sex, batch effects, complete blood cell count adjusted for in sensitivity analyses
Sun, et al., 2019, Circulation	Bogalusa Heart Study (BHS)	1485	BMI	DNA methylation β-value	Whole Blood	Age, sex, current smoking status cell composition (Houseman) included as fixed effects with batch array as a random effect
Wahl, et al., 2017, Nature	KORA, LOLIPOP, EPICORE	5458	DNA methylation β-value	BMI	Whole Blood	Top 20 PCs of control probes, cell composition (Houseman), age, gender, smoking status, physical activity index and alcohol consumption

545

546 **Table 2.** *Interaction between BMI and race/ethnicity in WHI between non-Hispanic whites and*

547 *African Americans.*

CpG Site	Effect Estimate	Standard Error	Z-score	P-value
cg25652701	-5.8E-04	1.4E-04	-4.29	1.76E-05
cg25212453	5.01E-04	1.6E-04	3.05	0.002
cg08122652	8.9E-04	3.6E-04	2.47	0.014
cg27113059	-2.3E-04	9.59E-05	-2.36	0.018
cg15391590	-2.4E-04	1.1E-04	-2.18	0.029

548

549

550 **Table 3.** *Predicting BMI from DNA methylation using elastic net regression*

Predictors	RMSE	Adjusted R^2	Median Absolute Deviation
398 CpG sites	0.07	0.32	0.040
398 CpG sites + Age	0.07	0.32	0.036
398 CpG sites + Age + Ethnicity	0.07	0.35	0.035
398 CpG sites + Age + Ethnicity + Cell Composition	0.07	0.35	0.035

551

552 **Table 4**. *Outliers in the prediction model compared to log-normalized cardiometabolic risk*

553 *factors. Model adjusted for race/ethnicity, smoking status, age and physical activity.*

	Estimate	SE	P-value
Waist Circumference (n=4356)			
High Epigenetic BMI	5.88E-03	4.07E-03	0.14
Low Epigenetic BMI	9.04E-04	3.98E-03	0.82
Blood Glucose (n=3823)			
High Epigenetic BMI	6.00E-02	5.80E-03	<2E-16
Low Epigenetic BMI	-1.86E-02	648E-03	0.004
Blood Triglycerides (n=3829)			
High Epigenetic BMI	7.63E-02	1.04E-02	2.64E-13
Low Epigenetic BMI	-6.34E-02	1.15E-02	4.41E-08
HDL-cholesterol (n=3832)			
High Epigenetic BMI	-3.75E-02	5.39E-03	4.55E-12
Low Epigenetic BMI	2.46E-02	5.99E-03	4.27E-05

LDL-cholesterol (n=3740)			
High Epigenetic BMI	-1.19E-02	6.05E-03	0.05
Low Epigenetic BMI	-239E-03	6.71E-03	0.72

554

555

556

557

558

559

560

561

562

563

564

A

B

C